A Comparative Study of Practitioner Perceptions of Selection Methods in the United Kingdom

Duncan J. R. Jackson
Birkbeck, University of London and University of Johannesburg

Chris Dewberry
Birkbeck, University of London

Jaroslava Gallagher
Arctic Shores Ltd

Liam Close
Birkbeck, University of London

Author Note

Duncan J. R. Jackson, Department of Organizational Psychology, Birkbeck, University of London and Faculty of Management, University of Johannesburg; Chris Dewberry, Department of Organizational Psychology, Birkbeck, University of London; Jaroslava Gallagher, Artic Shores Ltd, and Liam Close, Department of Organizational Psychology, Birkbeck, University of London.

The authors would like to thank George Michaelides and Tim Byne for their contributions to this article.

Correspondence concerning this article should be addressed to Duncan J. R. Jackson, Department of Organizational Psychology, Birkbeck, University of London, Clore Management Centre, Torrington Square, London, WC1E 7JL.
E-mail: dj.jackson@bbk.ac.uk

Abstract

In this study, 476 participants, divided into occupational psychology- (OP), Chartered Institute of Personnel and Development- (CIPD), human resource management- (HRM) qualified, and layperson subgroups, provided their perceptions of the validity, fairness, and frequency-of-use of employee selection methods. Results of a mixed-effects analysis of covariance revealed that respondent qualification background predicted the degree to which participant validity perceptions were aligned with research-based estimates of validity ($F[3, 29.39] = 20.06$, $p < .001$, $\eta^2 = .67$). Corrected pairwise comparisons suggested that perceptions of participants with CIPD and HRM backgrounds were not significantly more aligned with research estimates of validity than were the perceptions of laypeople. OP participant validity perceptions were significantly more aligned with research estimates than all other subgroups, ($p < .03$). Evidence was also found for some between-group consistency regarding frequency-of-use perceptions, but less between-group consistency was found vis-à-vis perceptions of fairness. Implications for decision making in employee selection are discussed.

Practitioner Points

- Knowledge about employee selection measures might not be effectively shared between respondents with CIPD- and HRM-related qualifications versus those with OP-related qualifications
- Laypeople and respondents with CIPD- and HRM-related qualifications were found to similarly deviate from up-to-date research findings about the validity of selection measures
- Respondents with OP-related qualifications were more closely aligned with up-to-date findings about the validity of selection measures than were other comparison groups

A Comparative Study of Practitioner Perceptions of Selection Methods in the United Kingdom

Effective employee selection is critical for the achievement of organisational strategic objectives (Bolander & Sandberg, 2013). Selection is a topic of increasing concern to organisations, particularly given recent economic volatility and the need to ensure that organisations sustain or improve their performance and reputations whilst remaining economical. Resource-based theory maintains that for organisations to hold competitive advantage, they need to have in their employment talent that is valuable and in short supply (e.g., Acedo, Barroso, & Galan, 2006). As organisations move towards a knowledge-based economy, fostering competitive advantage with respect to employee talent is becoming an increasing priority because "human capital" has been found to predict customer satisfaction, innovation, and the financial performance of organisations (Crook, Todd, Combs, Woehr, & Ketchen, 2011). While selection is a process that is important for an organisation's survival and performance, little is known about the knowledge that specific groups of practitioners have about the selection procedures they use to guide employment decisions. If practitioners do not hold the knowledge required to use selection procedures optimally, then the effectiveness of the selection process will be undermined. It is this issue, in the context of selection in the UK, upon which the current study is focussed.

Despite its importance to organisations, "existing research has usually paid little attention to how selection decision making takes place in real-life situations" (Bolander & Sandberg, 2013, p. 285). Psychometric tests, interviews, and other selection methods are routinely used to guide selection decisions in practice and a great deal of knowledge has accumulated on the development of employee selection methods and their psychometric characteristics. However, with the exception of assessment centres (ACs, e.g., Krause, Rossberger, Dowdeswell, Venter, & Joubert, 2011), relatively little is known about practitioner' perceptions of the psychometric or

other characteristics of selection methods (see Highhouse, 1997; Zysberg & Nevo, 2004). Employee selection methods (e.g., standardised tests) might represent "the greatest technological achievement in industrial and organizational (I-O) psychology over the past 100 years" (Highhouse, 2008, p. 333). But the effectiveness of selection methods, including those that are found to be reliable and valid, is likely to be dependent on the manner in which they are used and, therefore, the knowledge of practitioners.

The perceptions of practitioners and laypeople about selection methods are likely to play a role in their appropriate use (Highhouse, 2008; Rynes, Colbert, & Brown, 2002). For example, practitioner perceptions about the criterion-related validity and fairness of methods are likely to influence which methods are chosen and used by organisations. Additionally, layperson job applicant perceptions about the validity and fairness of selection methods are likely to influence the degree to which they respond positively and constructively to post-selection feedback. To date, a modestly sized literature database has accumulated on practitioner perceptions of employee selection procedures (e.g., Furnham, 2008; Hodgkinson & Daley, 1995; Rynes et al., 2002; Sanders, van Riemsdijk, & Groen, 2008). This literature has focused primarily on generalised practitioner perspectives on validity (see Highhouse, 2008) and applicant perspectives on fairness (e.g., Phillips & Gully, 2002). Although the discipline has made some progress towards developing an understanding of perceptions of selection methods, we are not aware of any published research that has attempted to directly compare the perceptions of specific practitioner subgroups against the perceptions of laypeople. We are also unaware of any research on the extent to which the perceptions of selection methods by these subgroups reflect evidence from published empirical research.

A search of the *Occupational Information Network* (O*NET) database reveals 503 occupations related to employee selection; many or most of which are likely to differ with respect to their qualification requirements and, thus, qualification-related background (e.g., human resource specialists, I-O psychologists, labour relations specialists, human resource managers, etc). Such differences in background could possibly be associated with systematic differences in perceptions of the validity and fairness of selection methods because the type of selection-related training that practitioners receive could depend on the qualifications that they attain.

The identification of qualification-dependent perceptual differences, particularly about the validity of employee selection procedures, could offer insights into how research findings are communicated within the community of practitioners. This is important for ensuring that all sectors of the practitioner community are informed about research developments in employee selection. Furthermore, by contrasting specific practitioner groups against layperson perceptions of selection methods, insights could be gained about how practitioners could focus their communication and guidance more effectively when communicating with laypeople on the topic of selection. This is important for helping to maintain a constructive dialogue with the wider community about a practice that affects all those either in or with the potential to be in employment.

In the present article, we offer a comparative study of the perceptions of the validity, fairness, and frequency-of-use of employee selection methods for respondents with qualifications related to occupational psychology[1] (OP), human resource management (HRM), and the Chartered Institute of Personnel and Development (CIPD[2]) in contrast to the perceptions of

[1] The UK term *occupational psychology* is equivalent to the US term *industrial-organizational psychology*.
[2] The CIPD is the main accreditation body for human resource practices in the UK.

laypeople. In particular, we aim to explore whether group-specific perceptions concur with evidence-based characteristics of selection methods, as documented in peer-reviewed, scholarly research.

Practitioner Perspectives on Validity

Studies on validity perceptions have generally reported a lack of concordance between research findings and practitioner perceptions. Almost three decades ago, Dakin and Armstrong (1989) found discrepancies between practitioner perceptions of validity and research-based estimates of validity. General mental ability (GMA), whilst typically ranked 1st in the research literature for validity (then and now, see Bertua, Anderson, & Salgado, 2005; Salgado et al., 2003; Schmidt & Hunter, 1998) was ranked 11th out of 12 methods by a practitioner group in New Zealand. The same group ranked work experience 1st, which, according to research estimates, typically ranks in the mid-to-low range for predictors in employment contexts (Quiñones, Ford, & Teachout, 1995).

Studies following Dakin and Armstrong (1989) in other countries also suggest discrepancies between perceptions and research estimates of validity. In the UK, Furnham (2008) found that ACs were ranked highest for perceived validity within a practitioner group, despite research findings suggesting otherwise (Arthur, Day, McNelly, & Edens, 2003; Gaugler, Rosenthal, Thornton, & Bentson, 1987; Schmidt & Hunter, 1998). Also in Furnham's study, personality tests were ranked 3rd out of 12 selection methods, even though some research suggests that personality test scores are among the weaker predictors of outcome performance (Judge & Zapata, 2015). Further discrepancies between perceptions and research have been revealed in the USA, where Terpstra (1996) found that a group of practitioners perceived *un*structured interviews to be the most valid method, despite findings suggesting that

unstructured interviews can potentially impair the quality of selection decisions (Kausel, Culbertson, & Madrid, 2016). Also in the USA, Rynes et al. (2002, p. 160) found "considerable discrepancies between research and practitioner beliefs" and, in a replication of this study in the Netherlands, Sanders et al. (2008, p. 1976) found "remarkable similarities" with Rynes et al. and, again, discrepancies between research findings and practitioner perceptions.

Highhouse (2008) suggests that, rather than necessarily indicating a lack of knowledge about selection methods, a lack of agreement between research and perceptions might imply that practitioners simply do not believe in the relevance of research findings to their practice. For example, although respondents in the Rynes et al. (2002) study tended to report that unstructured interviews were the most effective selection method, they also reported being aware of their documented pitfalls. Highhouse suggests that a possible reason for this might be misguided beliefs about selection methods. Notwithstanding these concerns, researchers have found evidence that the perceptions of practitioners who read the academic literature tend to be better-aligned with research findings than the perceptions of practitioners who do not read the academic literature (Rynes et al., 2002; Sanders et al., 2008). Also, it could be the case that information availability is dependent on the background of the practitioner and that misguided beliefs about selection could be alleviated with better-targeted or more "suitably-packaged" information about selection methods.

More than One Background, More than One Perspective?

The inconsistencies found between practitioner perceptions and research with respect to validity could be influenced by a lack of clear or accessible information about research findings (Campbell, 1990) and has possibly led to a pattern of results that are pervasive as they are of concern (Anderson, 2005). Practitioners who work in the context of employee selection derive

from more than one background. Previous studies have sampled participants with HRM-related backgrounds, including personnel consultants (Dakin & Armstrong, 1989), HRM practitioners (Furnham, 2008), human resource managers (Terpstra, 1996), and HRM professionals (Rynes et al., 2002; Sanders et al., 2008). However, no known study has attempted to contrast the perceptions of HRM practitioners against practitioners who have a background in other areas relevant to employee selection, such as OP. Also, those with HRM-related qualifications in the UK might either hold a degree in HRM, a degree in an HRM-related discipline, or might hold a professional qualification in HRM. Moreover, the CIPD offers a range of professional courses and qualifications for its members (see Farndale & Brewster, 2005) and, thus, an important distinction in the UK environment might be between CIPD qualifications and HRM-related qualifications.

Decision Making and Practitioner Background

The employee selection process is one that requires managers to engage in difficult and potentially risky decisions that could affect the ethical orientation and productivity of organizations, the well-being of participants, and the careers of job candidates. The *bounded rationality model* of decision making recognises that complete information is typically unavailable to decision makers because limited resources will often constrain the number of alternatives that can be evaluated (Puranam, Stieglitz, Osman, & Pillutla, 2015). This model also suggests that decision makers might have inadequate information available to them. Due to such constraints, decision makers might engage heuristics or "rules of thumb" to guide decisions, which might not necessarily have an empirical or a logical basis.

It is possible that, with respect to selection, information unavailability is more of a problem for some professional backgrounds than for others. Much of the research on selection

methods is published in journals with an OP orientation. For example, the most recent meta-analyses on personality, GMA, interviews, ACs, biodata, academic grades, and vocational interests were published in the *Journal of Occupational and Organizational Psychology*, the *European Journal of Work and Organizational Psychology,* the *Journal of Applied Psychology*, the *International Journal of Selection and Assessment*, and *Personnel Psychology* (Arthur et al., 2003; Huffcutt, Culbertson, & Weyhrauch, 2014; Roth, BeVier, Switzer, & Schippmann, 1996; Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990; Salgado et al., 2003; Salgado, Anderson, & Tauriz, 2015; Salgado & Táuriz, 2014; Van Iddekinge, Roth, Putka, & Lanivich, 2011). Arguably, these publications are oriented more towards readers with an OP background and might not be accessed as readily by the HRM audience as they are by the OP audience. This possibly exacerbates a knowledge-gap for HRM practitioners with respect to selection methods.

Across multiple HRM-related topics, both Rynes et al. (2002) and Sanders et al. (2008) found staffing to be *the* area where the greatest divergences were evident between practitioner perceptions about best practice and results from research findings. But an implied assumption in the literature appears to be that these divergences are true for practitioners from *any* background working in an HRM-related area. If it were identified that divergences were true for practitioners with a background related to HRM but not for practitioners with a background in OP, then such findings could inform on bounded rationality and information availability as they relate to decision making in the selection context. Specifically, it could suggest that information availability might depend on barriers to information-sharing that arise because of the professional background of the decision maker. This could also lead to the targeting of interventions for assisting practitioners with different backgrounds to share knowledge more effectively.

This leads to our first Research Question:

Research Question 1: In contrast to the views of laypeople, does the degree of divergence between perceptions of the validity of selection methods and up-to-date research-based evidence depend on whether a practitioner is from a CIPD, HRM, or OP background?

For best practice to be guided by research in employee selection, it seems reasonable to expect that, in the very least, CIPD, HRM, and OP perceptions of the validity of employee selection methods would align more closely to research estimates than the perceptions of laypeople. Furthermore, the layperson comparison group potentially provides insights into how practitioners might better communicate information about selection methods to the wider community.

Our focus in Research Question 1 is on up-to-date research evidence, implying our interest in a familiarity with contemporary literature. The literature on the validity of selection measures has developed and changed over the years (e.g., Schmidt & Hunter, 1998; Schmidt, Oh, & Shaffer, 2013) and, thus, it is important that practitioners who use measures for the purposes of employee selection are aware of the latest research developments in the area.

Practitioner Perspectives on Fairness and Frequency-of-Use

In addition to views on validity, of interest to the research and practice community are perceptions about the fairness of selection methods and how these perceptions relate to views on the frequency with which particular methods are used. This literature generally refers to or implies a conception of fairness that aligns with the definition of procedural justice: i.e., the

perceived equity of rules and decisions relating to a procedure (e.g., Slocum & Hellriegel, 2009). We also follow this definition of fairness in the current paper.

Most of the literature to date on fairness perceptions has centred on the perspective of job applicants (see Anderson, Salgado, & Hülsheger, 2010; Ryan & Ployhart, 2000). Applicant perceptions of fairness have been investigated in a diverse range of geographic locations, including Spain, Portugal, the Netherlands, the United States, Greece, Singapore, France, and India (Anderson & Witvliet, 2008; Gilliland, 1993; Moscoso & Salgado, 2004; Nikolaou & Judge, 2007; Phillips & Gully, 2002; Snyder & Shahani- Denning, 2012; Steiner & Gilliland, 1996). This research body tends to suggest that, regardless of geographic location, the more popular methods among applicants include interviews, work samples, and résumés. In contrast, less popular methods include graphology, personal contacts, and integrity tests.

Since the seminal work of Steiner and Gilliland (1996), research on applicant fairness perspectives has provided an important message to organisations wishing to balance potential candidate anxiety against validity evidence. However, also important are practitioner perceptions about fairness and how these views relate to perceptions about the frequency with which selection methods are used. Do practitioners with CIPD, HRM, or OP qualifications differ with respect to their views on perceptions of use-frequency and fairness regarding selection methods? If differences arise, then there is a possibility that job candidates might be exposed to a particular set of selection procedures, depending on practitioner background. If members of a specific subgroup perceive that a selection method is both popular and fair, then they might be more inclined to apply that approach. Also, within practitioner sub-groups, do rankings of use-frequency agree with rankings of fairness? If not, the perception might be that the methods used frequently in practice are not necessarily those that are also perceived to be fair.

For practitioners, a discrepancy of this type could result in concerns about the state of practice. For laypeople, it could represent a source of dissatisfaction with selection procedures. This leads to the following Research Questions:

Research Question 2: Do differences exist between CIPD, HRM, and OP practitioners and laypeople on perceptions of the frequency with which selection methods are used?

Research Question 3: Do differences exist between CIPD, HRM, and OP practitioners and laypeople on perceptions of the fairness of selection methods?

Research Question 4: Within each of CIPD, HRM, and OP, and layperson groups, do perceptions of frequency-of-use agree with perceptions of the fairness of selection procedures?

As a supplementary research aim, within each subgroup, we will also test for the extent to which frequency-of-use perceptions relate to validity perceptions for CIPD, HRM, and OP practitioners, and laypeople. If any discrepancies arise between perceptions of use-frequency and validity, then this might suggest a perceived divide between how selection decisions are typically made in practice and how they possibly "should" be made (based on the participants' understanding of the validity of selection methods).

Method

Participants

A total of 476 respondents from the UK participated in this study, consisting of 287 females and 188 males with a mean age of 40.04 (*SD* = 12.55). Participants were invited to complete an online survey that was posted on the *LinkedIn* networking website and shared with professional groups in the United Kingdom, including the CIPD, the *Association for Business Psychology*, the *British Psychological Society Division of Occupational Psychology*, and the *Psychometrics Forum*. In order to gain perceptions from laypeople, the survey was also distributed to general participants in the UK via the respondent facility provided by the *Survey Monkey* organization. Detailed demographic information on participants is provided in Table 1.

Materials and Procedure

For comparison purposes, participants were grouped by background according to demographic data relating to their qualification area (see Table 1 for details). This classification resulted in four participant subgroups: laypeople and CIPD, HRM, and OP groups. Other demographics relevant to the present study were collected, including vocational and educational information (see Table 1).

Participants were required to rank 13 employee selection methods according to their perceptions about validity, fairness, and frequency-of-use. Because laypeople might be unfamiliar with the term "validity" in the context of employee selection, participants were asked to rank order selection methods based on their perceived "effectiveness" for guiding employment decisions. For the benefit of participants who might be more familiar with it, the term "validity" was also explicitly provided in the questionnaire as part of the definition of "effectiveness". We only refer to the term "validity" in this article for ease of communication with our readership. Note also that our use of the term "effectiveness" as a proxy for "validity" is not without precedent in the current research context (see Rynes et al., 2002; and Sanders et al., 2008).

In the ranking task, we were constrained to including only methods that were likely to have some familiarity to laypeople. We did not include integrity or situational judgment tests because of uncertainties about whether laypeople would have knowledge about them. Also, we did not distinguish between structured and unstructured interviews because this distinction assumes prior knowledge about interview type. We did not include work samples because preliminary research into our respondents revealed that they did not typically (or ever) use work sample tests in their practice. Moreover, we used the general term "personality" because we could not assume knowledge of particular personality models. In terms of our research-based rankings for personality, we used estimates related to conscientiousness. We used the simplified term "training" as opposed to introducing a distinction between the point and behavioural consistency methods of evaluating training and experience (see Schmidt & Hunter, 1998). Here, we also wanted to avoid conceptually confounding training-related assessments with assessments of work experience. We used point method research-based estimates of validity because the point method represents a fairly simple approach to rating training-related behaviour and, moreover, it seemed unlikely that laypeople would be aware of the more technically-comprehensive behavioural consistency approach (see Ash, 1983).

Analysis

Ranked validity perceptions. Mean ranks of selection method validity perceptions by qualification area subgroup were contrasted against contemporary research-based rankings of selection methods. The research perspective was based on a search of the PsycINFO database with a view to finding the most recent meta-analytic results available for each selection method (as listed in Table 2). To summarize areas where subgroup perceptions differed from research-based estimates, we calculated the total number of discrepancies between mean ranks for

perceived validities and ranked estimates from published research. To allow for a margin of error, we also calculated the number of discrepancies between perceived validity and research-based validity estimates that were within |3| rankings of each other. This interval was based on an analogue of the *standard error of measurement* rounded up to the nearest integer based on an average agreement index (Spearman's $\bar{r}_S$, discussed below) and average *SD*s for ranks across all groups (Gatewood, Feild, & Barrick, 2016; Howell, 2007). Differences in ranking patterns across subgroups were tested using the Kruskal-Wallis test (see Siegel & Castellan, 1988). We applied standard Bonferroni corrections (see Field, 2013) to our Kruskal-Wallis tests, given the number of tests involved (resulting in a corrected alpha criterion value of .004).

For within-group comparisons, we applied the Wilcoxon signed ranks test because all within-group tests in the present study were concerned with pairwise comparisons between rankings. We also estimated interrater agreement within each subgroup using Kendall's coefficient of concordance (Kendall's *W*), which is applicable to ranked data (Siegel & Castellan, 1988). While no agreed guidelines have been developed for acceptable levels of Kendall's *W*, the statistic is interpreted on a scale from 0 (complete disagreement) to 1 (complete agreement). We also estimated the average Spearman correlations between respondent rankings (Spearman's $\bar{r}_S$). The Spearman's $\bar{r}_S$ statistic is a conversion of Kendall's *W* into more familiar, correlation-based terms (Howell, 2007).

Mixed-effects analysis of covariance and Cohen's *d*. To provide a clearer, statistically controlled perspective on the validity perceptions data, we applied a mixed-effects analysis of covariance (ANCOVA, see Field, 2013; Howell, 2007). Initially, this involved generating deviation scores for each individual by taking the research-based ranking away from the perceived ranking for each selection method listed in the above ranking task. In turn, we took

the sum of these deviations for each selection method to create overall deviation scores for each individual. Higher deviation scores indicate less general alignment with research estimates and lower scores indicate greater general alignment with research estimates. Overall deviation scores were used as outcome values in the mixed-effects ANCOVA, where (a) age and gender were treated as covariates, (b) qualification level (other, high school, undergraduate degree, and postgraduate degree) was treated as a fixed factor, and (c) qualification area (layperson, CIPD, HRM, and OP) was treated as a random factor[3]. With respect to the covariates in this model, age has been associated with experience and, thus, a potentially enhanced capability to apply a set of decision rules (Bruine de Bruin, Parker, & Fischhoff, 2012). Also, some evidence exists for gender differences in certain recall tasks (Voyer, Postma, Brake, & Imperato-McGinley, 2007; Yarmey, Jacob, & Porter, 2002). Because of these findings, age and gender were thought to require statistical control to aid the interpretability of our results.

Of primary interest, with respect to the ANCOVA model, was an indication of the extent to which participant qualification area explained variance in overall deviation scores whilst controlling for the effects of qualification level, age, and gender. Our preliminary foci were the statistical significance of each effect in the model and the effect sizes as estimated by partial η^2. We also investigated specific between-group comparisons with respect to variance explained in deviation scores. To this end, we used pairwise comparisons and tested for group differences using Bonferroni- and Sidak-adjusted estimates for multiple comparisons (see Field, 2013; Tabachnick & Fidell, 2001). We also estimated between-group differences using Cohen's *d*, which provides a useful description for comparisons between subgroup pairs.

[3] We treated qualification area as a random effect because there are numerous other, potentially viable qualification areas (e.g., general management, labour relations, international business, business administration, general administration, and social psychology) that could have been submitted for analysis. In fact, the O*NET database identifies a total of 502 occupation areas that relate to employee selection.

Ranked frequency-of-use and fairness perceptions. On frequency-of-use and fairness perceptions, we presented mean ranks for each qualification area subgroup and tested between-group differences using Kruskal-Wallis tests. We also tested within-group comparisons using Wilcoxon signed ranks tests. Of interest was whether within-group perceptions of the frequency with which selection methods are applied aligned with perceptions of the fairness of selection methods for guiding employment decisions. Thus, any significant within-group differences would likely suggest that frequency-of-use perceptions for a given method do not necessarily align with the perception that the method in question is a fairer approach to evaluating individuals for employment purposes. All significance tests were adjusted for multiple comparisons in the same manner described above for validity perceptions. For both the perceived frequency-of-use and perceived fairness ranks, we also estimated interrater agreement within each qualification area subgroup using Kendall's W and Spearman's $\bar{r}_S$.

All analyses in this study were conducted using SPSS (version 23) except for the estimation of Cohen's *d* and confidence intervals for Cohen's *d*, which were conducted using the R package *effsize* (Torchiano, 2016).

Results

Validity Perceptions

Table 2 shows a descriptive analysis of the rank ordering of validity perceptions, by background, of selection methods in relation to research-based estimates. In Table 2 there is an indication of a difference between the response patterns of laypeople in comparison to other groups with relevant training, given that, within a margin of |3| ranks, 6 discrepancies with research estimates were found with laypeople and between 3 and 4 discrepancies were found with the remaining groups with employee-selection-relevant background training. In terms of

specific discrepancies between subgroups and contemporary research estimates, significant and notable differences were observed on GMA, ACs, work experience, reference checks, personality, and years of education. GMA and ACs were ranked higher by OP respondents than by other groups. Work experience was ranked lower by OP than by other groups. Reference checks were ranked higher by laypeople than by OP respondents. Ranks for personality were variable across groups and years of education was ranked lower by OP respondents. Interrater agreement was also estimated for validity rankings and was slightly higher for OP respondents (W = .49) than for other groups. As might be expected, agreement among laypeople was the lowest among the respondents (W = .29).

Omnibus results from the mixed effects ANCOVA suggested a significant and large partial main effect for qualification area, $F(3, 29.39) = 20.06$, $p < .001$, $\eta^2 = .67$. All other effects were non-significant ($p > .16$) and very small ($\eta^2 < .03$). This suggests that among those modelled, the sole factor contributing meaningfully to the prediction of deviation scores was the qualification background of respondents.

Pairwise comparisons with Bonferroni or Sidak adjustments[4] revealed that differences between respondents with a background in OP contrasted against all other groups were significant ($p < .03$). All other group differences were non-significant, including differences between those with a CIPD background and laypeople ($p = .59$), an HRM background and laypeople ($p = .09$), and a CIPD and HRM background ($p > .99$). To help illustrate the magnitude of these pairwise differences, we plotted marginal mean deviation scores for the qualification area factor, adjusted for the covariates age and gender (see Figure 1). The marginal mean value for laypeople was the highest among the groups sampled. Those with CIPD and HRM qualifications returned similar marginal means, which were slightly, although not

[4] Bonferroni and Sidak corrections returned almost identical results.

significantly, lower than the mean for laypeople. The lowest marginal mean, which was significantly lower than the other groups, was that associated with the OP group.

Table 3 shows Cohen's *d* values for each pair of subgroups in this study, the results of which were similar to those suggested in the ANCOVA model. Standardised differences between the discrepancy scores of laypeople and those with CIPD and HRM backgrounds were small. In contrast, the difference between layperson and OP respondent discrepancy scores were large.

In summary, and with reference to *Research Question 1*, our results suggest that when overall discrepancies between perceptions and research-based estimates of validity were considered, the degree of divergence between perceptions and research findings depended on qualification background. Specifically, perceptions of validity from respondents with CIPD or HRM qualifications tended to be about as divergent from contemporary research estimates as were the perceptions of laypeople. Conversely, those with OP qualifications tended to be in closer alignment with research findings than all other groups sampled.

Frequency-of-Use and Fairness Perceptions

Table 4 shows mean ranks by qualification area for frequency-of-use and fairness perceptions. Regarding frequency-of-use, results were fairly consistent across all groups (see *Research Question 2*). Interviews were, on average, unanimously ranked as the most frequently used method, followed by work experience, and then reference checks. Interests and age were generally ranked lowest. As a supplement to this analysis, we tested for within-group paired comparisons with respect to validity perceptions and perceptions of frequency-of-use (see Appendix, Table A1). Cross-referencing Tables A1, 2, and 3, of interest were the findings with respect to the OP group for interviews (ranked 4th for validity, yet 1st for frequency-of-use, *p*

< .001) and generally for reference checks (ranked 6th or lower for validity, yet 3rd for frequency-of-use, $p < .01$). This suggests that some discrepancies exist between what is found to be valid in research versus the frequency with which methods are used in practice.

Relative to frequency-of-use perceptions, somewhat less between-subgroup consistency was found with respect to perceptions of fairness (see Table 4 and *Research Question 3*). The OP group tended to rank job try-outs 1st in terms of fairness to job applicants. Whereas CIPD respondents ranked interviews 1st and HRM respondents and laypeople ranked work experience 1st. The OP group ranked GMA tests 2nd for fairness versus the CIPD group, who ranked work experience 2nd, and the HRM and layperson groups, who ranked interviews 2nd. There was greater between-group consistency regarding the lower fairness rankings with all groups tending to agree that age ranked lowest and all but the layperson group tending to agree that interests ranked 2nd-lowest.

Multiple, significant within-group differences shown in Table 4 suggest that perceptions of use frequency did not necessarily align with perceptions of fairness for selection methods for each group (see *Research Question 4*). This was particularly so for the layperson group, who returned 11 significant differences between frequency and fairness ranks. However, even those with CIPD, HRM, and OP training returned between 7 and 8 significant differences, suggesting that discrepancies between frequency-of-use and fairness largely occurred regardless of qualification background.

Discussion

We sought to contribute to the literature on perceptions about the validity, frequency-of-use, and fairness of employee selection methods. Given that employee selection is central to the effective functioning of organisations (Crook et al., 2011), the perceptions of practitioners,

particularly with respect to the validity of selection methods, could represent a key determinant in the appropriate use of such methods. Relevant previous research has hitherto focused on perceptions from individuals primarily with HRM backgrounds (Dakin & Armstrong, 1989; Furnham, 2008; Rynes et al., 2002; Sanders et al., 2008; Terpstra, 1996). We sought to expand on previous work in this area by differentiating between individuals with CIPD, HRM, and OP qualifications in contrast to laypeople with no training specifically relevant to employee selection practices.

Validity Perceptions versus Research Evidence: An OP versus HRM/CIPD Divide?

Regarding *Research Question 1*, we found evidence to suggest that qualification area was an important factor for determining the extent to which perceptions of the validity of selection methods aligned with up-to-date validity estimates from empirical research. Results of a mixed-effects ANCOVA revealed that, in terms of their perceptions of the validity of a set of 13 selection methods, responses of CIPD- ($p = .59$) and HRM-qualified ($p = .09$) participants were not significantly more closely aligned to contemporary research estimates than were the perceptions of laypeople. In contrast, perceptions of those with a background in OP were significantly more aligned with current research estimates than all groups under scrutiny ($p < .03$). In terms of effect size, Figure 1 suggests that the results for OP were more closely aligned with research estimates than those for other groups. Estimates of effect size based on partial η^2 values revealed that the main effect for qualification area explained around 67% of the variance in scores that summarised the discrepancy between perceptions and research-based validity estimates ($F[3, 29.39] = 20.06$, $p < .001$, $\eta^2 = .67$). A large portion of this effect was due to the presence of OP participants, as evidenced by the fact that when OP participants were removed from the analysis, the effect size for qualification area decreased by around 52% (F[2,

73.12] = 6.53, $p < .01$, $\eta^2 = .15$). Effects from the perspective of Cohen's *d* showed similar results with only OP respondents showing substantially greater alignment with research-based validity estimates than laypeople (see Table 3).

Our findings suggest that discrepancies between perceptions of the validity of selection methods versus up-to-date research estimates might be more extreme for practitioners with CIPD or HRM qualifications than for practitioners with qualifications in OP. These findings are important for three reasons. Firstly, if practitioners with CIPD or HRM qualifications do not typically use research evidence to guide their practice, then an important element of the research enterprise has been overlooked by a sizable practitioner group. Research, particularly on the topic of the validity of selection methods, should assist practitioners to refine and improve their practice. Secondly, and related to the first point, selection decisions are among perhaps the most important high-stakes decisions in which practitioners in this area are involved. Employee selection decisions have implications for individuals and their careers, they can affect individual well-being, and can also give rise to resource-intensive litigation for organizations (Konstam, Celen-Demirtas, Tomek, & Sweeney, 2015; Williams, Schaffer, & Ellis, 2013). Therefore, ensuring a connection between research and practice for HR-qualified individuals is a priority.

The judicious application of employee selection methods by trained professionals is, from the perspective of litigation and with reference to professional and ethical conduct, of concern: particularly in the light of research on applicant perceptions of justice related to research based estimates of operational validity. Anderson et al. (2010) found a number of significant and strong correlations between job applicant perceptions of justice dimensions and validity estimates (particularly perceptions of favourability, scientific evidence, providing opportunities to perform, and face validity). Thus, what job applicants perceive as fair might be aligned to

some degree with what is found to be valid with relation to selection methods. This means that professionals who apply selection methods could find themselves doing so with a group of applicants who are at least somewhat accurate in their perceptions of selection methodology. Such professionals should, therefore, work to ensure that they are up-to-date with research estimates of validity for reasons of professional integrity as well as for ethical reasons. Our results raise concerns for the CIPD- and HRM-trained groups in relation to the layperson referent.

Thirdly, our results suggest potential restrictions in the sharing of what might have inadvertently become OP-specific knowledge with CIPD- or HRM-qualified individuals. Such restrictions could result from the fact that validity results, vis-à-vis employee selection methods, are often published in OP-related journals (e.g., Arthur et al., 2003; Huffcutt et al., 2014; Schmidt & Hunter, 1998; Van Iddekinge et al., 2011). Restrictions in the cross-discipline sharing of information might also be compounded by a division between CIPD-, HRM-, and OP-specific professional and academic events such as conferences (e.g., those organized by the *Institute of Work Psychology* and the *Division of Occupational Psychology* versus the CIPD in the UK; and those organized by the *Society for Industrial and Organizational Psychology* versus the *Society for Human Resource Management* in the USA). Might these divisions promote the development of knowledge silos? If so, then knowledge sharing between sub-disciplines could suffer as a result. However, as Highhouse (2008) suggests, it is possible that the research findings are, in fact, communicated and understood by practitioners but are not seen as being relevant to them. If this is true, then perhaps our findings suggest that research outcomes need to be "repackaged" for HR-qualified individuals in such a manner that makes their relevance clearer to that audience. Such efforts would still require knowledge-sharing between WP and

HR professionals and, therefore, steps to increase current levels of communication and discussion between them.

Our findings possibly inform on the *bounded rationality* model of decision making as it pertains to employee selection. The bounded rationality model suggests that decision makers often have inadequate or incomplete information available to guide their decisions and, accordingly, might arrive at decisions using heuristics rather than data (Puranam et al., 2015). Our findings suggest that information availability in employee selection might depend to some extent on qualification area, with some areas perhaps having better access to relevant information than others. As detailed in our introduction, recent meta-analyses for a range of popular selection methods have appeared in OP-related journals. It seems logical to expect that those with OP backgrounds would be more likely to have access to information printed in OP journals.

Our general, descriptive finding of discrepancies between practitioner perceptions and research-based estimates vis-à-vis validity bears similarities with previous research findings (e.g., Rynes et al., 2002; Sanders et al., 2008). However, as is the case in other studies on this topic, perceived rankings relating to specific selection methods suggested cross-study differences. The slightly lower (3rd) ranking of GMA by CIPD and HRM practitioners in our study was not as low as the ranking of 11th/12 in Dakin and Armstrong (1989). Our findings also differed from those of Furnham (2008), in that CIPD and HRM-qualified respondents only ranked ACs 5th on average in our study, but ACs were ranked 1st/12 in Furnham's study. Furthermore, CIPD and HRM respondents ranked personality 8th and 9th, respectively, in our study. However, in Furnham's study, personality was ranked 3rd. It is not possible to match these rankings perfectly because of differences in both study methods and the number and type of selection techniques

presented to respondents across studies. Also, the benchmarks for validity rankings change over the years as the research area develops: hence our focus on up-to-date validity estimates.

Another potential reason for cross-study variation could be that there tends to be middling-to-low agreement within subgroups of respondents with respect to perceptions of validity. Previous research in this area has not, to our knowledge, assessed agreement relating to perceptions of validity, which we regard as a potential oversight. Table 2 shows estimates of interrater agreement for ranked validity perceptions (see Kendall's *W* in Table 2), none of which was particularly high. Agreement among laypeople was lowest (*W* = .29), as might be expected. Agreement among OP-qualified respondents was highest (*W* = .49). These findings suggest the possibility of general uncertainty among our participants about validity rankings. Thus, mean ranking discrepancies relating to specific selection methods for laypeople and HR-qualified respondents are probably not particularly important or, for that matter, generalisable across different studies. Possibly more important than discrepancies relating to specific methods is the overall perspective on discrepancies between research estimates and perceptions (i.e., that offered by our mixed-effective ANCOVA and our estimates of Cohen's *d*). This level of analysis could be interpreted as an indication of overall discrepancies between perceptions of validity and research-based estimates. Rather than focusing on discrepancies relating to specific methods, as has been the case in several previous studies, we recommend that future research, instead, models a representation of overall discrepancies.

Frequency-of-Use and Fairness

Table 4 shows mean ranks for perceptions regarding the frequency-of-use and fairness of selection methods (see *Research Question 2*). On average, we found evidence for between-group consistencies regarding perceptions about how frequently particular methods are used.

Interviews were ranked 1st for frequency of use, work experience 2nd, and reference checks 3rd by all three groups. Also, interests and age were generally ranked lowest for use-frequency. Whilst frequency-of-use was associated with some between-group consistency (i.e., 12/13 of the between-groups tests resulted in non-significant differences), perceptions of fairness were less consistent between groups see *Research Question 3*). With the exception of age, interests, and grades (ranked lowest across groups), there was almost no consensus between groups with respect to mean rankings of fairness.

With respect to perceptions of frequency-of-use and relatively higher between-group consistency, it is possible that respondent judgments were based on experience and direct observation. On fairness perceptions and the lack of between-group consistency, it is possible that respondent judgments were based on beliefs about fairness and even beliefs about the definition of fairness. While the literature on the legality of selection procedures might be more clear-cut (see Furnham, 2008; Snyder & Shahani- Denning, 2012; Terpstra, Kethley, Foley, & Limpaphayom, 2000; Terpstra & Rozell, 1997), the issue of fairness, per se, is open to interpretation and what might be considered "legally-safe" territory might not necessarily be considered fair in this context. Perhaps the very concept of fairness needs clarification in the discipline as it applies to employee selection decisions and this could represent an avenue for future research. Although a great deal is known about fairness perceptions as they apply to job applicants (e.g., Anderson & Witvliet, 2008), far less is known about practitioners' perceptions of fairness.

Bearing in mind within- and between-group inconsistencies, particularly with respect to fairness perceptions, we also found some evidence to suggest that perceptions of frequency-of-use did not necessarily align with perceptions of fairness (see the *within-groups* tests in Table 4

and *Research Question 3*). These discrepancies were particularly evident for the layperson group, which could raise concerns about perceptions by non-practitioners about how selection procedures are used in practice (i.e., the belief that practitioners do not necessarily commit methods believed to be "fair" to frequent use). We also tested for differences between frequency-of-use perceptions and validity perceptions within each group and found multiple differences, particularly within the layperson and the OP group (cross-reference Tables A1, 2, and 4). This suggests that laypeople and OP practitioners might perceive discrepancies between the validity of a given method and the frequency with which that method is applied.

Limitations and Future Research Directions

Ours is a single study carried out in the UK and further studies that separate different practitioner groups are necessary to ensure the generalisation of our results. It would be of particular interest to the community to know if practitioner group dependencies, regarding perceptions of selection methods, exist in different countries. Although job applicant perceptions of selection methods often generalize with surprising consistency across different countries (e.g., Anderson & Witvliet, 2008; Gilliland, 1993; Snyder & Shahani- Denning, 2012), it is uncertain as to whether the patterns that we observed for practitioners will generalise to other regions.

The reader should also be aware that response rates for our study are not known. This is because we sought to maximise the possibility of responses from a potentially difficult-to-access group of practitioners by using an internet-based survey. The trade-off here was that it is not possible to provide an accurate estimate of the population *N* because (a) it is not possible to precisely define our sample frame due to our use of a general internet-based invite to participants and (b) the exact sample frame is also difficult to specify since the levels in our *qualification background* factor were, themselves, sampled from a population of potentially relevant

qualification backgrounds. While the O*NET database specifies 503 occupations relevant to employee selection, it is difficult or impossible to define a precise sample frame for background profiles relevant to each of these occupations.

Our study is limited by the selection methods that we listed in our ranking task. Our choice of methods precluded us from making direct comparisons between practitioner judgments of fairness against such judgments made by job applicants in the published literature. However, these constraints occurred because we needed to take into consideration the layperson perspective in our study. We assumed that laypeople did not have any training in an HRM-related discipline and, therefore, would not hold knowledge needed to decipher the nuances associated with particular selection methods. Future studies, perhaps benefitting from the patterns that we have identified here, could include a sample of only those with specialist HRM-related knowledge.

On the point of practitioner qualifications, our study contrasted CIPD-, HRM-, and OP-qualified respondents against a referent layperson perspective. It would be of interest to broaden the scope of the present study and to see if different within-group effects emerge when a larger set of practitioner groups are included for analysis. There were also a small proportion of participants in our study who reported holding CIPD- ($N = 3$) or HRM-related ($N = 5$) qualifications but who only held high school-level education. While this might be expected for CIPD-qualified participants because CIPD qualifications are professional qualifications, this is perhaps less expected for HRM-qualified individuals. As a precaution, we re-ran our analyses with the relevant HRM-qualified participants omitted and our results were almost identical to what they were with the inclusion of these participants.

On our use of ANCOVA, concerns might be raised about some of the restrictive assumptions associated with this technique and whether, indeed, those assumptions were met. A visual inspection of the histogram associated with our discrepancy score outcome variable suggested that it approximated a normal distribution. We tested for the assumption of homogeneity of regression slopes (i.e., involving the covariates age and gender and the factors qualification level and qualification area). No significant relationships were found in this respect, except for the relationship between age and qualification area. However, the magnitude of this relationship was small, $F(3, 451) = 7.52$, $p < .001$, $\eta^2 = .05$. Moreover, re-running the model with the age covariate removed resulted in almost the same results (i.e., no significant predictors except for qualification area, $F[3, 25.57] = 19.48$, $p < .001$, $\eta^2 = .70$) as when it was included in the model. In addition, we tested for the homogeneity of variance assumption using Levene's test, which, of reassurance, yielded a non-significant result ($F[13, 452] = 901$, $p = .553$). We also assessed Cook's distances for the model, all of which were < 1, in keeping with expectations. Moreover, the proportion of standardised residuals ≥ |1.96| and ≥ |2.58| was 4.94% and 1.07% respectively, as is expected in general linear models (Field, Miles, & Field, 2012).

Concluding Comments

The findings of our study suggest that, in terms of their perceptions of the validity of selection methods, CIPD- and HRM-qualified respondents were not significantly more aligned with contemporary research estimates than were laypeople. The perceptions of OP respondents were, in contrast, significantly more closely aligned with research estimates. These results are suggestive of (a) a possible breakdown in the sharing of employee-selection-relevant knowledge with CIPD- and HRM-qualified practitioners or (b) a perception by CIPD- and HRM-qualified practitioners that employee selection research lacks relevance to practice. This implies that

either more effort needs to be made to represent selection validity research in publications that are read by practitioners or that validity research needs to be presented in such a manner that highlights its relevance to practitioners. Without such knowledge-sharing, it could be that the applied discipline is not fully benefitting from research-based guidance. We also found evidence that perceptions of the frequency with which selection methods are used did not necessarily match with perceptions of fairness. Moreover, our results suggested general uncertainty about the relative fairness of selection methods. In the light of our findings, perhaps further studies are needed that compare both applicant and practitioner views on validity and fairness and how such perspectives might influence how organizations choose and apply selection methods.

References

Acedo, F. J., Barroso, C., & Galan, J. L. (2006). The resource-based theory: dissemination and main trends. *Strategic Management Journal, 27*, 621-636. doi: 10.1002/smj.532

Anderson, N. (2005). Relationships between practice and research in personnel selection: Does the left hand know what the right is doing? In A. Evers, N. Anderson & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 1-24). Malden, MA: Blackwell.

Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment, 18*, 291-304. doi: 10.1111/j.1468-2389.2010.00512.x

Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment, 16*, 1-13. doi: 10.1111/j.1468-2389.2008.00404.x

Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154. doi: 10.1111/j.1744-6570.2003.tb00146.x

Ash, R. A. (1983). The behavioral consistency method of training and experience evaluation: Content validity issues and completion rate problems. *Public Personnel Management, 12*, 115-127.

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology, 78*, 387-409. doi: 10.1348/096317905X26994

Bolander, P., & Sandberg, J. (2013). How employee selection decisions are made in practice. *Organization Studies, 34*, 285-311. doi: 10.1177/0170840612464757

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2012). Explaining adult age differences in decision-making competence. *Journal of Behavioral Decision Making, 25*, 352-360. doi: 10.1002/bdm.712

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 687–732). Palo Alto, CA: Consulting Psychologists Press.

Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J., Jr. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology, 96*, 443-456. doi: 10.1037/a0022147

Dakin, S., & Armstrong, S. (1989). Predicting job performance: A comparison of expert opinion and research findings. *International Journal of Forecasting, 5*, 187-194.

Farndale, E., & Brewster, C. (2005). In search of legitimacy: personnel management associations worldwide. *Human Resource Management Journal, 15*, 33-48.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Thousand Oaks: Sage.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks: Sage.

Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment, 16*, 300-305. doi: 10.1111/j.1468-2389.2008.00436.x

Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2016). *Human resource selection* (8th ed.). Mason, OH: Cengage Learning.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493-511. doi: 10.1037/0021-9010.72.3.493

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*, 694-734. doi: 10.5465/AMR.1993.9402210155

Highhouse, S. (1997). Understanding and improving job-finalist choice: The relevance of behavioral decision research. *Human Resource Management Review, 7*, 449-470. doi: 10.1016/S1053-4822(97)90029-2

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 333-342. doi: 10.1111/j.1754-9434.2008.00058.x

Hodgkinson, G. P., & Daley, N. (1995). Knowledge of, and attitudes towards, the demographic time bomb. *International Journal of Manpower, 16*, 59-76. doi: 10.1108/01437729510100794

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Bemont, CA: Thomson Wadsworth.

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment, 22*, 297-309. doi: 10.1111/ijsa.12078

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98. doi: 10.1037/0033-2909.96.1.72

Judge, T. A., & Zapata, C. P. (2015). The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the big five personality traits in predicting job performance. *Academy of Management Journal, 58*, 1149-1179. doi: 10.5465/amj.2010.0837

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes, 137*, 27-44. doi: 10.1016/j.obhdp.2016.07.005

Konstam, V., Celen-Demirtas, S., Tomek, S., & Sweeney, K. (2015). Career adaptability and subjective well-being in unemployed emerging adults. *Journal of Career Development, 42*, 463-477. doi: 10.1177/0894845315575151

Krause, D. E., Rossberger, R. J., Dowdeswell, K., Venter, N., & Joubert, T. (2011). Assessment center practices in South Africa. *International Journal of Selection and Assessment, 19*, 262-275. doi: DOI 10.1111/j.1468-2389.2011.00555.x

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, 283-314.

Moscoso, S., & Salgado, J. F. (2004). Fairness reactions to personnel selection techniques in Spain and Portugal. *International Journal of Selection and Assessment, 12*, 187-196. doi: 10.1111/j.0965-075X.2004.00273.x

Ng, T. W. H., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology, 93*, 392-423. doi: 10.1037/0021-9010.93.2.392

Nikolaou, I., & Judge, T. A. (2007). Fairness reactions to personnel selection techniques in Greece: The role of core self-evaluations. *International Journal of Selection and Assessment, 15*, 206-219. doi: 10.1111/j.1468-2389.2007.00382.x

Phillips, J. M., & Gully, S. M. (2002). Fairness reactions to personnel selection techniques in Singapore and the United States. *The International Journal of Human Resource Management, 13*, 1186-1205. doi: 10.1080/09585190210149475

Puranam, P., Stieglitz, N., Osman, M., & Pillutla, M. M. (2015). Modelling bounded rationality in organizations: Progress and prospects. *The Academy of Management Annals, 9*, 337-392. doi: 10.1080/19416520.2015.1024498

Quiñones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology, 48*, 887-910.

Romano, J., Kromrey, J. D., Coraggio, J., & Skowronek, J. (2006). *Appropriate statistics for ordinal level data: Should we really be using t-test and cohen's d for evaluating group differences on the NSSE and other surveys?* Paper presented at the Florida Association of Institutional Research, Cocoa Beach, Florida.

Roth, P. L., BeVier, C. A., Switzer, F. S., III, & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81*, 548-556. doi: 10.1037/0021-9010.81.5.548

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75*, 175-184. doi: 10.1037/0021-9010.75.2.175

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26*, 565-606.

Rynes, S. L., Colbert, A. E., & Brown, K. C. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*, 149. doi: 10.1002/hrm.10029

Salgado, J. F., & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology, 12*, 3-30.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rollan, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology, 88*, 1068-1081. doi: 10.1037/0021-9010.88.6.1068

Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology, 88*, 797-834. doi: 10.1111/joop.12098

Salgado, J. F., & Táuriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology, 23*, 3-30. doi: 10.1080/1359432X.2012.716198

Sanders, K., van Riemsdijk, M., & Groen, B. (2008). The gap between research and practice: A replication study on the HR professionals' beliefs about effective human resource practices. *International Journal of Human Resource Management, 19*, 1976-1988. doi: 10.1080/09585190802324304

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmidt, F. L., Oh, I., & Shaffer, J. (2013, November). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 95 years of research findings.* Paper presented at the Personnel Testing Council Metropolitan Washington (PTCMW), Washington DC.

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Slocum, J. W., & Hellriegel, D. (2009). *Principles of organizational behavior* (12th ed.). Mason, OH: South-Western.

Snyder, J., & Shahani- Denning, C. (2012). Fairness reactions to personnel selection methods: A look at professionals in Mumbai, India. *International Journal of Selection and Assessment, 20*, 297-307. doi: 10.1111/j.1468-2389.2012.00601.x

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81*, 134-141. doi: 10.1037/0021-9010.81.2.134

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Pearson.

Terpstra, D. E. (1996, May). The search for effective methods. *HR Focus, 73,* 16.

Terpstra, D. E., Kethley, R. B., Foley, R. T., & Limpaphayom, W. (2000). The nature of litigation surrounding five screening devices. *Public Personnel Management, 29*, 43.

Terpstra, D. E., & Rozell, E. J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management, 26*, 483.

Torchiano, M. (2016). effsize: Efficient Effect Size Computation. (Version 0.7.0) [R package]. https://CRAN.R-project.org/package=effsize.

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology, 96*, 1167-1194. doi: 10.1037/a0024343

Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic Bulletin and Review, 14*, 23-38. doi: 10.3758/BF03194024

Williams, K. Z., Schaffer, M. M., & Ellis, L. E. (2013). Legal risk in selection: An analysis of processes and tools. *Journal of Business and Psychology, 28*, 401-410. doi: 10.1007/s10869-013-9299-4

Yarmey, A. D., Jacob, J., & Porter, A. (2002). Person recall in field settings. *Journal of Applied Social Psychology, 32*, 2354-2367. doi: 10.1111/j.1559-1816.2002.tb01866.x

Zysberg, L., & Nevo, B. (2004). "The smarts that counts?": Psychologists' decision-making in personnel selection. *Journal of Business and Psychology, 19*, 117-124. doi: 10.1023/B:JOBU.0000040275.16133.15

Table 1
Frequencies for Sample Demographics

Demographic information	Classification				Total
	LP	CIPD	HRM	OP	
Position					
HR director	1	6	9	2	18
HR regional manager	0	1	1	0	2
HR senior manager	0	4	4	2	10
HR manager	1	7	13	5	26
HR assistant manager	2	4	6	2	14
HR associate	3	6	4	2	15
HR assistant	4	6	7	4	21
HR consultant	2	14	22	26	64
Other HR position	10	15	7	14	46
Non-HR employment	181	0	0	0	181
Job applicant	25	0	0	0	25
Full-time student	19	0	0	0	19
Unemployed	35	0	0	0	35
Total	283	63	73	57	476
Highest qualification level					
Postgraduate degree	117	39	41	46	243
Undergraduate degree	85	19	17	11	132
High School	49	3	5	0	58
Other	36	2	5	0	43
Qualification area					
CIPD	0	63	0	0	63
HRM	0	0	33	0	33
OB	0	0	24	0	24
HRD	0	0	11	0	11
CMC	0	0	5	0	5
OP	0	0	0	57	57
Other or none	283	0	0	0	283
Ethnicity					
Black	17	2	6	1	26
White	235	53	53	48	389
Eastern and Greater European	10	4	1	6	21
Indian	6	3	5	1	15
Other or non-response	15	1	8	1	25
Gender					
Female	170	42	44	31	287
Male	113	20	29	26	188
Non-response	0	1	0	0	1

Note. LP = layperson/no human resource-related qualifications; CIPD = Chartered Institute of Personnel Management; HRM = human resource management; OP = occupational psychology; OB = organizational behaviour; HRD = human resource development; CMC = career management and coaching.

Table 2

Descriptive Research-Based Validity Estimates of Selection Methods versus Perceptions of Validity by Qualification Area

Method	Research Estimate		Laypeople		CIPD		HRM		OP		*p*
	r	Rank	M_{rank}	Rank	M_{rank}	Rank	M_{rank}	Rank	M_{rank}	Rank	
GMA tests	.53[a]	1	5.53	4	4.86	3	4.88	3	2.91	1	**<.01**
Interviews	.46[b]	2	4.52	2	3.51	2	4.29	2	4.95	4	.01
Job try-outs	.44[c]	3	5.05	3	4.97	4	4.89	4	4.09	3	.33
Assessment centres	.36[d]	4	6.82	7	5.62	5	5.85	5	3.84	2	**<.01**
Biodata	.35[e]	5	7.27	8	7.84	10	6.56	6	7.25	7	.45
Academic grades	.32[f]	6	8.27	10	8.56	11	8.30	11	8.82	11	.54
Work experience	.27[g]	7	3.64	1	3.19	1	3.88	1	5.37	5	**<.01**
Reference checks	.26[c]	8	6.33	6	7.29	7	7.10	8	8.25	9	**<.01**
Personality	.21[h]	9	8.56	11	7.76	8	7.66	9	6.42	6	**<.01**
Interests	.14[i]	10	9.65	12	10.08	12	9.95	12	9.88	12	<.68
Training	.11[j]	11	7.34	9	7.81	9	7.86	10	7.49	8	<.34
Years of education	.10[c]	12	6.25	5	6.76	6	7.08	7	8.51	10	**<.01**
Age	.03[k]	13	10.88	13	11.76	13	11.41	13	12.39	13	**<.01**
Discrepancies (Abs)	-	-	-	10	-	11	-	10	-	10	-
Discrepancies (|3|)	-	-	-	6	-	4	-	3	-	3	-
Kendall's *W*	-	-	.29		.42		.33		.49		-
Spearman's $\bar{r}_S$	-	-	.28		.41		.32		.48		-

Note. [a]from Salgado et al. (2003) for medium-level job complexity (GMA = general mental ability). Bertua et al. (2005) and Salgado and Anderson (2003) found operational validity estimates in a similar range (.48 and .52 respectively) for GMA, which do not affect the ranking of GMA in this study. [b]From Huffcutt et al. (2014) for Level 2 structured interviews; [c]from Hunter and Hunter (1984); [d]from Arthur et al. (2003); [e]from Rothstein et al. (1990); [f]from the overall coefficient in Roth et al. (1996); [g]from Quiñones et al. (1995); [h]based on the operational validity estimate for general forced choice conscientiousness measures from Salgado and Táuriz (2014). Salgado et al. (2015) also present meta-analytic estimates for conscientiousness for specific occupational groups, but not a general estimate across forced-choice formats. [i]from Van Iddekinge et al. (2011); [j]from McDaniel, Schmidt, and Hunter (1988) for the training and experience point method; [k]from Ng and Feldman (2008) for supervisor ratings corrected for interrater reliability; *p* = *p*-values for mean differences in ranks between laypeople, participants with CIPD (Chartered Institute of Personnel and Development), human resource management (HRM) and occupational psychology (OP) qualifications based on Kruskal-Wallis tests. Values in bold indicate significant differences after applying a standard Bonferroni correction for 13 tests (corrected alpha = .004). In all sub-samples, Kendall's *W* was significant to $p < .01$. Spearman's $\bar{r}_S$ = average Spearman's correlation between judges' rankings, based on Kendall's *W*. Discrepancies (Abs) and Discrepancies (|3|) = number of absolute differences and differences within |3| ranks, respectively, between rankings based on research versus rankings based on perceptions out of a total of 13 possible differences. The interval of |3| was based on an analogue of the *standard error of measurement* rounded up to the nearest integer, based on average Spearman's $\bar{r}_S$ and average SDs for ranks across all groups.

Table 3
Standardised Between-Group Discrepancy-Score Comparisons

Comparison	M_1	SD_1	M_2	SD_2	*d*	95% CI (*d*)		Description[a]
LP(M_1) – CIPD(M_2)	45.58	8.90	41.49	7.00	- 0.47	-0.75,	- 0.20	Small
LP(M_1) – HRM(M_2)	45.58	8.90	41.48	10.31	- 0.45	-0.71,	- 0.18	Small
LP(M_1) – OP(M_2)	45.58	8.90	35.51	8.96	1.13	0.83,	1.42	Large
CIPD(M_1) – HRM(M_2)	41.49	7.00	41.48	10.31	< 0.01	-0.34,	0.34	Negligible
CIPD(M_1) – OP(M_2)	41.49	7.00	35.51	8.96	0.75	0.37,	1.12	Medium
HRM(M_1) – OP(M_2)	41.48	10.31	35.51	8.96	0.61	0.25,	0.97	Medium

Note. Lower mean scores indicate closer alignment with research findings. LP = laypeople; CIPD = Chartered Institute of Personnel Development-qualified; HRM = human resource management-qualified; OP = occupational psychology-qualified. *d* = Cohen's *d*; CI = confidence interval. [a]Based on thresholds and descriptions presented in Romano, Kromrey, Coraggio, and Skowronek (2006) and in Torchiano (2016). We also estimated standardised differences using Hedges' *g*: however, these estimates did not alter the conclusions suggested in the results presented above.

Table 4

Descriptive Estimates for Perceptions of Frequency-of-Use and Perceived Fairness of Selection Methods

	Frequency-of-Use Between-Groups					Perceived Fairness Between-Groups					Within-Groups (p^b)			
Method	LP	CIPD	HRM	OP	p^a	LP	CIPD	HRM	OP	p^a	LP	CIPD	HRM	OP
GMA	(5)6.64	(5)6.11	(6)6.62	(4)4.55	.11	(4)5.53	(5)5.52	(3)4.80	(2)3.92	<.01	**<.01**	.12	**<.01**	**<.01**
Intvw	(1)2.51	(1)2.16	(1)2.08	(1)1.81	.39	(2)4.50	(1)3.68	(2)4.79	(5)5.29	.02	**<.01**	**<.01**	**<.01**	**<.01**
Jtout	(8)7.49	(7)7.65	(7)7.61	(5)6.29	.98	(3)4.68	(3)3.98	(4)4.85	(1)3.54	.03	**<.01**	**<.01**	**<.01**	**<.01**
ACs	(10)8.32	(9)7.89	(9)7.87	(8)7.74	.28	(7)7.07	(4)5.48	(5)6.04	(3)4.06	**<.01**	**<.01**	**<.01**	**<.01**	**<.01**
Bdata	(6)7.27	(8)7.84	(5)6.56	(7)7.26	.26	(9)8.17	(11)9.05	(8)7.51	(7)8.40	.07	**<.01**	<.01	.03	.06
Grade	(7)7.48	(6)7.25	(8)7.80	(11)9.00	.56	(10)8.23	(10)8.59	(10)8.31	(11)8.98	.31	**<.01**	**<.01**	.11	.02
Wkexp	(2)3.67	(2)3.22	(2)3.68	(2)3.77	.46	(1)3.66	(2)3.87	(1)3.83	(4)4.79	**<.01**	.90	.09	.46	.02
Rfchk	(3)4.73	(3)4.00	(3)4.65	(3)4.48	.14	(5)5.83	(6)6.70	(7)7.00	(8)8.23	**<.01**	**<.01**	**<.01**	**<.01**	**<.01**
Persy	(11)9.32	(11)9.06	(11)8.63	(9)7.97	.01	(12)9.16	(9)8.14	(11)8.46	(6)6.29	**<.01**	.28	.10	.65	.03
Intrst	(13)10.19	(12)10.57	(12)10.28	(12)11.19	.51	(11)9.08	(12)9.14	(12)9.10	(12)9.23	.81	**<.01**	**<.01**	**<.01**	.16
Train	(9)8.12	(10)8.21	(10)8.01	(10)8.90	<.01	(8)7.35	(7)7.33	(9)7.58	(7)7.62	.78	**<.01**	.06	.26	**<.01**
Edu	(4)5.31	(4)5.81	(4)6.03	(6)6.35	.06	(6)6.48	(8)7.43	(6)6.69	(9)8.40	**<.01**	**<.01**	**<.01**	.06	**<.01**
Age	(12)9.94	(13)11.22	(13)11.17	(13)11.68	**<.01**	(13)11.27	(13)12.08	(13)12.04	(13)12.25	**<.01**	**<.01**	**<.01**	**<.01**	**<.01**
W	.37	.48	.42	.42		.31	.40	.33	.44		-	-	-	-
$\bar{r}_S$	.37	.47	.41	.40		.31	.39	.32	.43		-	-	-	-

Note. GMA = general mental ability test; Intvw = interviews; Jtout = job try outs; ACs = assessment centers; Bdata = biodata; Grade = academic grades; Wkexp = work experience; Rfchk = reference checks; Persy = personality; Intrst = interests; Train = training; Edu = years of education. The method column is ordered from highest to lowest research-based validity estimates; mean ranks appear in parentheses; [a]*p*-values for mean differences in ranks between laypeople (LP), participants with Chartered Institute of Personnel Development (CIPD) qualifications, human resource management (HRM) qualifications, or occupational psychology (OP) qualifications based on Kruskal-Wallis tests. [b]*p*-values based on Wilcoxon signed ranks test for paired comparisons between frequency of use and perceived fairness within each sub-sample. Values in bold indicate a significant difference after applying a standard Bonferroni correction for 13 tests (where corrected alpha = .004). *W* = Kendall's *W*. In all sub-samples, Kendall's *W* was significant to $p < .01$. $\bar{r}_S$ = average Spearman's correlation between judges' rankings, based on Kendall's *W*.

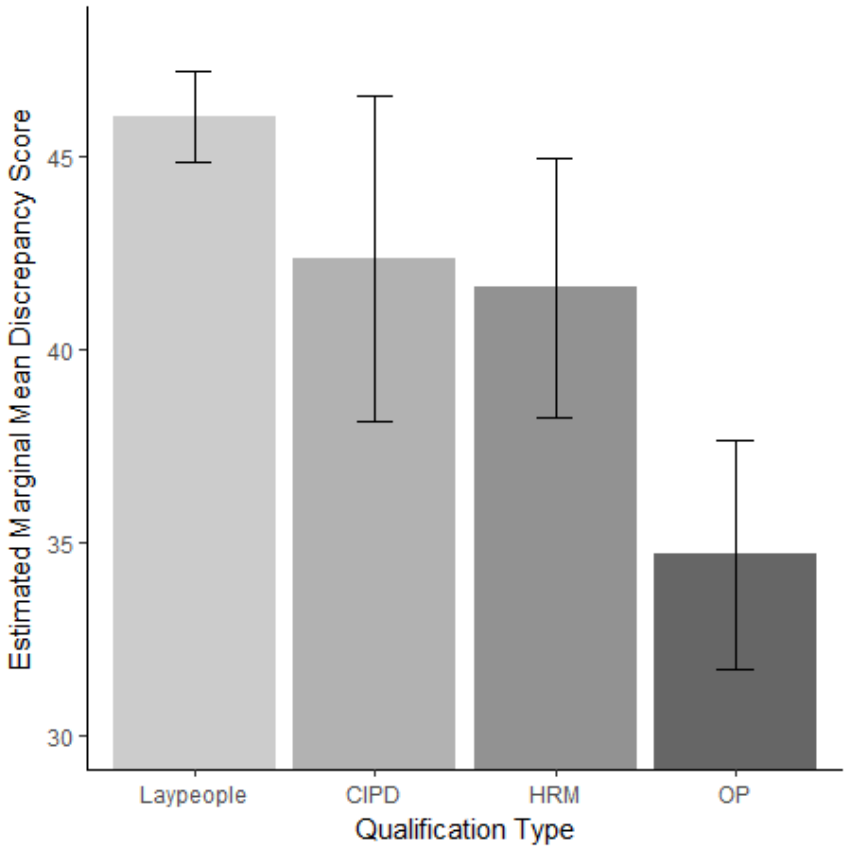

Figure 1. Bar chart showing estimated marginal mean discrepancy scores from an analysis of covariance, corrected for the covariates age and gender, as a function of respondent qualification area (Laypeople, CIPD = Chartered Institute of Personnel and Development, HRM = human resource management, OP = occupational psychology). Discrepancy scores reflect overall differences between perceptions of validity versus research-based validity estimates summed for 13 selection methods: thus, lower scores indicate closer alignment with research estimates. Bonferroni- and Sidak-corrected differences between scores from OP-qualified respondents and all remaining subgroups were significant ($p < .03$). All other differences were non-significant.

Appendix

Table A1

Within-Group Paired Comparisons for Validity and Frequency-of-Use Perceptions of Employee Selection Methods

Method	LP	CIPD	HRM	OP
GMA	**<.01**	**<.01**	**<.01**	**<.01**
Interviews	**<.01**	**<.01**	**<.01**	**<.01**
Job try-outs	**<.01**	**<.01**	**<.01**	**<.01**
Assessment centres	**<.01**	**<.01**	**<.01**	**<.01**
Biodata	**<.01**	.02	.01	.20
Academic grades	**<.01**	**<.01**	.04	.01
Work experience	.36	.97	.53	**<.01**
Reference checks	**<.01**	**<.01**	**<.01**	**<.01**
Personality	**<.01**	.03	.02	**<.01**
Interests	<.01	.20	.27	.24
Training and experience	**<.01**	.38	.47	**<.01**
Years of education	**<.01**	<.01	<.01	**<.01**
Age	**<.01**	.08	.51	**<.01**

Note. LP = laypeople; CIPD = Chartered Institute of Personnel Development-qualified; HRM = human resource management-qualified; OP = occupational psychology-qualified. All *p*-values based on Wilcoxon signed ranks tests for paired comparisons (i.e., comparisons between perceptions of validity and frequency of use within each group). Values in bold indicate a significant difference after applying a standard Bonferroni correction for 13 tests (where corrected alpha = .004).